Structural Breaks and Growth Regimes in India's UPI Ecosystem: A Piecewise Regression Analysis of Policy Push, Demonetisation, and Pandemic Effects

Kaju Nath
Department of MHSS, National Institute of Technology Agartala, Jirania-799046, Tripura, India
Suman Das*
Department of Education, National Institute of Technology Calicut, Kozhikode-673601, Kerala, India
Amit Bikram Chowdhury, Nabendu Debnath, Rajib Mallik
Department of MHSS, National Institute of Technology Agartala, Jirania-799046, Tripura, India

Uncorrected proof

Abstract

India's Unified Payments Interface (UPI) scaled from zero to 22,347 million monthly transactions between April 2016 and April 2026, with a decade-cumulative transaction value of approximately ₹1,099 lakh crore. Using NPCI's monthly series, this paper examines three candidate structural breaks in UPI's growth path, coinciding with Demonetisation (November 2016), COVID-19 onset (March 2020), and the post-lockdown recovery (July 2020), through a sequence of log-linear, structural-dummy, piecewise, and quadratic trend models. Because these dates are drawn from known policy events rather than detected statistically, we treat the resulting break tests with caution and report a supplementary search that does not fully corroborate them, leaving a confirmation-bias risk we do not consider resolved. The piecewise specification fits the data substantially better than a log-linear benchmark, though its residuals show meaningful serial correlation that widens the true uncertainty around individual coefficients; we address this with heteroskedasticity- and autocorrelation-robust standard errors throughout. Rather than treating the estimated post-COVID shift as a fixed multiplier, we compare the fitted growth trajectories directly and find the gap is largest immediately after the shock and narrows substantially thereafter, a pattern we describe as a substantial but non-permanent change in trajectory. A quadratic specification points to decelerating growth with a turning point around April 2024, consistent with, though not sufficient to establish, S-curve-style maturation. We interpret these patterns as temporal associations rather than causal effects, and discuss the resulting identification, serial-correlation, and generalisability limitations for comparable digital payment programmes elsewhere.

JEL Classification: E42, G28, L86, O33

Keywords: UPI; digital payments; structural break; piecewise regression; Digital India.

1. Introduction

The Unified Payments Interface (UPI) is one of the most significant adoption stories in the history of retail payments in India. The platform, launched by the National Payments Corporation of India (NPCI) in April 2016, made its first non-zero month in July 2016 when it handled 0.09 million transactions. By April 2026, monthly volume had reached 22,347 million transactions, an annualised run-rate of around 268 billion transactions valued at approximately ₹348 lakh crore per year. Over the decade the platform processed around 751.9 billion transactions worth a cumulative ₹1,099 lakh crore. These figures are verified against the authors' source NPCI dataset. This scale places UPI among the largest retail payment systems globally by transaction count, though cross-network comparisons of this kind are sensitive to definitional and reporting differences and are not the focus of the empirical analysis that follows.

Three developments are commonly discussed as candidate drivers of this growth: the Digital India policy architecture, comprising the Jan Dhan financial-inclusion programme, the Aadhaar biometric identity system, and the interoperable JAM Trinity of platforms, which built supply-side capacity for digital payments; the November 2016 Demonetisation, which withdrew 86 percent of currency in circulation overnight and created acute short-term demand for digital alternatives; and the COVID-19 pandemic, which from mid-2020 coincided with a shift toward contactless payment behaviour and rapid growth in the digital payment user base. We treat these as candidate, overlapping explanations rather than as established causal mechanisms: the time-series methods used in this paper can identify when the growth path changed and by how much, but cannot on their own separate the distinct contribution of each factor, particularly because several supply-side developments, including Jio's 4G rollout (September 2016) and the BHIM app launch (December 2016), occurred within the same narrow window as Demonetisation.

These narratives are not mutually exclusive, but they carry different policy implications. A Demonetisation-first account suggests that a forced fiscal shock can trigger a digital transition. A COVID-first account suggests that an exogenous public-health shock can accelerate a fintech ecosystem toward large-scale activation. A Digital India account emphasises patient, sustained investment on the supply side ahead of any shock. This paper's empirical contribution is to document, as precisely as the data allow, the timing, size, and duration of the identified structural breaks, and to discuss, with appropriate caveats about causal identification, how each candidate explanation aligns with the estimated pattern.

We take up this question using NPCI's published monthly series spanning 121 months (April 2016-April 2026). The three break dates used throughout, November 2016, March 2020, and July 2020, are drawn from known policy and public-health events rather than detected through a data-driven search procedure; this is standard practice in the Chow-test tradition where external events motivate candidate dates (Chow, 1960), but it also means the exercise cannot on its own rule out confirmation bias. We report a rolling Chow search over all candidate months as a supplementary diagnostic (Section 3.3, Section 4.3); as detailed there, this search does not resolve the confirmation-bias concern, since its global maximum does not coincide with our three hypothesised dates, and we treat an unrestricted multiple-break search (Section 4.3) as a

Uncorrected Proof

further, only partially reassuring, check. Our approach includes: (1) a log-linear trend model as a null baseline; (2) a structural-dummy model with level shifts at Demonetisation and Post-COVID; (3) a piecewise slopes-and-level-shifts specification allowing distinct growth rates across five periods; and (4) a quadratic log-trend model. We report Chow structural-break tests, a rolling Chow search, the CUSUM parameter-stability test, and Durbin-Watson/Ljung-Box autocorrelation diagnostics; the latter reveal meaningful residual serial correlation in the piecewise model, a limitation we address directly with HAC-consistent standard errors and an attempted AR(1)-corrected sensitivity check, which converged numerically but produced economically uninterpretable coefficients (Section 4.2); we treat robustness to serial correlation as an open question rather than resolved in either direction.

The remainder of the article is organised as follows. Section 2 reviews the post-2020 literature on digital public infrastructure, pandemic-driven adoption, and structural-break modelling of payment systems. Section 3 details the data and econometric methodology, including the limitations of the a priori break-date approach. Section 4 presents empirical results for all four models with diagnostic testing. Section 5 examines the annual growth trajectory across financial years. Section 6 discusses financial-inclusion implications and their limits. Section 7 presents a comparative interpretive assessment, framed as an interpretive synthesis rather than a formally identified decomposition. Section 8 addresses growth maturation and projections. Section 9 presents policy implications, and Section 10 concludes with an explicit discussion of limitations.

2. Literature Review

This section is organised as a cumulative argument rather than a catalogue of studies. It develops three linked claims that structure the empirical design. First, because a payment rail is a network good whose value rises with the number of users on each side of the market (Katz & Shapiro, 1985; Rochet & Tirole, 2003), adoption should be path-dependent and prone to coordination failure, so that a large shock can move the system from a low to a higher adoption path, with the size of the move depending on the infrastructure already in place (Crouzet et al., 2023; Higgins, 2024). Second, diffusion theory implies that any such acceleration should be followed by decelerating proportional growth as the pool of potential users is depleted (Griliches, 1957; Bass, 1969; Rogers, 2003). Third, because the shocks relevant to the Indian case are few, overlapping and dated from external events, the econometric literature on structural change shows what break tests can detect and what they cannot (Perron, 1989; Andrews, 1993; Bai & Perron, 2003; Hansen, 2001). Sections 2.1 to 2.4 take these claims in turn, and Section 2.5 draws them together into the expectations that Models M1 to M4 are used to examine.

2.1 Digital Public Infrastructure and Open Payment Platforms

The theoretical underpinning of UPI's design draws on platform economics and the theory of two-sided markets (Rochet & Tirole, 2003) and on the economics of network externalities (Katz & Shapiro, 1985), in which the value to a user of joining a payment network depends on how

Uncorrected proof

many others have joined. Open, interoperable payment infrastructure can lower transaction costs, reduce proprietary lock-in, and allow competitive entry by third-party providers on a shared rail. Within this framework, Parker and Van Alstyne (2005) characterise the conditions under which open platforms outperform proprietary ones in generating network effects, a framework commonly applied to NPCI's architecture; whether it applies with equal force here is discussed rather than assumed in Section 7. The implication for the present paper is that interoperability matters less as a design feature in itself than as a means of lowering the coordination cost of adoption, which is the channel through which the shocks discussed in Sections 2.2 and 2.3 are expected to operate.

Bank participation on UPI grew from 21 to 713 institutions over the sample period, an expansion consistent with an open, interoperable platform design in which government sets the rules while private firms compete on user experience; we treat this as background context rather than as evidence our own models estimate. The World Bank (2023) identifies interoperable digital identity, payment, and data-sharing systems as important components of digital public infrastructure for advancing financial inclusion and productivity. India's JAM architecture provides a prominent example of this broader DPI approach. Evidence on the Indian stack is consistent with this reading. Alonso et al. (2023) describe how identity, payment and data-exchange layers built as shared components allowed private innovation to build on a common core, and Muralidharan, Niehaus and Sukhtankar (2016) show that biometrically authenticated payment delivery in an Indian welfare programme reduced leakage and made payments faster and more predictable, evidence that identity and payment layers lower the cost of moving people from cash to electronic channels. Evidence from outside India points the same way: Jack and Suri (2014) find that Kenyan households with access to the M-PESA mobile-money network were better able to maintain consumption after income shocks, which suggests that payment rails can matter for welfare beyond transaction convenience.

Sahay et al. (2020) develop cross-country measures of digital financial inclusion and show that digital finance is associated with greater financial inclusion and economic growth. They further argue that access to digital infrastructure and well-integrated electronic systems became particularly important during the COVID-19 crisis. Their empirical evidence, however, largely predates the pandemic and therefore does not establish the persistence of post-COVID adoption gains. Whether their cross-country argument extends to the within-India timing documented in Section 4 is a further question this paper's single-country time series cannot resolve on its own.

Taken together, these studies establish that open, identity-linked payment rails reduce adoption frictions and can raise welfare. They are, however, largely cross-sectional or programme-specific, and they do not offer a time-series account of when and by how much an adoption path change, which is the gap this paper addresses (Section 2.5).

2.2 Demonetisation: Evidence on Digital Payment Effects

Unconfected Proof

The demonetisation literature supports two distinct claims that are often run together. The first is that the cash shortage raised the relative attractiveness of electronic payments in the short run; the second is that the resulting adoption proved durable. Chodorow-Reich et al. (2020) and Agarwal et al. (2024) speak mainly to the first, using geographic and transaction-level variation in exposure to the shock. Crouzet et al. (2023), using merchant-level data from a large electronic-wallet provider, speak more directly to the second: they find that the temporary cash shortage produced a lasting rise in retailers' adoption of electronic payments, and interpret it as the shock acting as a coordination device in a technology with network externalities. Their result matters for this paper because it gives a theoretical reason to expect a demonetisation-era change in the growth path that does not fully reverse, which is the pattern Model M3 is used to examine.

The literature on demonetisation's payment effects is substantial. Chodorow-Reich et al. (2020) exploit geographic variation in the severity of India's demonetisation and find that districts experiencing more severe cash shortages showed reductions in economic activity and faster adoption of alternative payment technologies. Their results provide credible evidence that demonetisation accelerated substitution toward non-cash payment methods, although the study does not provide the user-income or demographic decomposition examined indirectly in Section 6.

Agarwal, Ghosh, Li and Ruan (2024) study transaction-level Indian consumer spending around the 2016 demonetisation and its relationship to digital-payment adoption. What our own data show directly (Section 4.4) is that volumes neither fully reverted after the demonetisation window closed nor sustained a rate above the organic steady-era growth that follows, consistent with, though not establishing on its own, the specific pattern this paper's Model M3 estimates for the same window: a negative level shift at the demonetisation boundary alongside an elevated within-era growth slope (Section 4.2).

An important counterpoint to the demonetisation-as-catalyst narrative is that several supply-side developments overlapped with the same window: Jio's 4G rollout (September 2016), the BHIM app launch (December 2016), and merchant QR-code subsidies. We are not aware of a peer-reviewed study that separately identifies each of these developments' contribution to digital-payment growth using a credible identification strategy. Because these developments fall within the same narrow window as demonetisation in the present data, our models cannot separate their distinct contributions regardless of what the wider literature finds; this is a genuine identification limitation of the demonetisation-era estimates, which we flag here and return to in Section 3 and Section 7 rather than resolve in favour of any single account.

In sum, the demonetisation evidence supports an adoption response with at least partial persistence, but these studies rely on district-level, transaction-level or merchant-level variation rather than on the national UPI series analysed here, and, as noted above, we are not aware of work that separates the shock from the contemporaneous supply-side launches. For this reason

Uncorrected proof

Section 3 treats the Demonetisation window as a maintained hypothesis rather than as an identified effect.

2.3 COVID-19 as Structural Accelerant: Post-2020 Evidence

The pandemic literature builds on the same coordination logic. Auer, Cornelli and Frost (2020) discuss how the pandemic raised the appeal of contactless and digital payment channels and what this may imply for the future of cash, and Sahay et al. (2020), discussed in Section 2.1, argue that digital infrastructure became particularly valuable during the crisis. Read through the coordination lens of Section 2.1, a pandemic shock should accelerate adoption most where acceptance networks and user habits are already in place, which is the digital-readiness complementarity discussed next.

The post-2020 literature offers evidence broadly consistent with COVID-19 reducing practical and behavioural barriers to digital payments rather than creating demand from nothing, though this is one reading among several. It is plausible that economies with more advanced pre-pandemic digital infrastructure would experience faster, larger, and more durable adoption shifts during a shock like COVID-19, a "digital readiness" complementarity; if such a cross-country pattern holds within India as well, it would be consistent with the post-COVID level shift documented in Section 4 building on an already-embedded platform, though our single-country series cannot test this complementarity directly and we are not aware of a verified cross-country study we can cite in support of it.

Merchant-side acceptance and QR-code diffusion may have contributed to post-pandemic UPI growth, but the aggregate NPCI series used here does not allow us to identify that mechanism separately. Higgins (2024) analyses how network externalities between consumers and merchants shape the adoption of cashless payments in Mexico, an emerging-market analogue of the merchant-acceptance channel that would matter for UPI as well, although it is outside the scope of the present aggregate series.

2.4 Structural Break Modelling of Payment Systems

The econometric methodology we employ draws on the structural-break and regime-switching literature in time-series analysis. Chow's (1960) test for parameter stability remains the standard approach for testing pre-specified break dates, appropriate when theory or, as here, known external events identify candidate break points; it is not designed to discover break dates from the data itself. Bai and Perron (2003) extend the framework to unknown, multiple break dates and note that financial time series frequently exhibit several structural breaks; we adopt their general motivation for a multi-regime specification while relying, unlike Bai-Perron, on externally specified rather than statistically detected break dates, a choice with the confirmation-bias risk discussed in Section 3.1.

Uncorrected proof

A further strand of this literature concerns what such tests can and cannot show. Perron (1989) demonstrates that ignoring a structural break biases standard unit-root tests towards failing to reject a unit root, which is directly relevant to the sensitivity of the ADF results reported in Section 3.3. Andrews (1993) shows that when the break date is unknown, the maximum of a sequence of Chow-type statistics does not follow the standard F distribution and requires its own critical values, so the rolling search in Section 4.3 is used here descriptively, to locate where instability is largest, rather than as a formal test; Hansen (2001) surveys these developments. Where residuals are serially correlated, as in the models estimated here, Newey and West (1987) heteroskedasticity- and autocorrelation-consistent covariance estimation is the standard remedy for inference, although it does not correct misspecification of the break structure itself.

Model selection via information criteria such as AIC is one way to formally compare piecewise and smooth trend specifications, though such comparisons are themselves sensitive to residual autocorrelation, a caveat this paper addresses directly in Sections 3.3 and 4.5. Each of the structural-break and regime-switching methods reviewed here assumes, at minimum, that residuals are approximately serially uncorrelated and homoskedastic once the break structure is accounted for; violations of either assumption inflate the apparent precision of coefficient estimates and can bias information-criteria-based model selection toward more heavily parameterised specifications. As Section 4.5 documents, the piecewise model estimated here shows pronounced residual autocorrelation, a limitation common in monthly macro-financial series with strong trend components and one we address through explicit caveats on inference rather than a fully corrected GLS or ARMA specification, which we flag as a direction for future work. Nor is the evidence in this broader literature uniformly supportive of large, durable shock effects: the demonetisation-era simultaneity concern discussed above (Section 2.2) is one reason to question whether demonetisation had an independent effect once contemporaneous supply-side developments are accounted for, and apparent pandemic-era payment gains more broadly can partly reflect base effects and reporting changes rather than a durable behavioural shift. We treat net-effect estimates in this literature as contested rather than settled, and frame our own findings accordingly.

Diffusion-of-innovations frameworks (Griliches, 1957; Bass, 1969; Rogers, 2003) suggest that decelerating proportional growth in a quadratic log-trend specification like our M4 can be consistent with movement toward later stages of adoption; we return to this interpretation, and its limits, in Section 8. In these models adoption follows an S-shaped path in which the adoption rate rises, peaks and then declines as the pool of potential adopters is depleted, so deceleration is informative about saturation only to the extent that the potential market is bounded.

A recurring policy question concerns the potential tension between broadening reach toward low-income users, reflected in declining average transaction values, and the commercial sustainability models of payment service providers. Both issues are relevant to projections beyond our sample period.

uncorrected proof

2.5 Synthesis and Analytical Framework

The preceding sections can be drawn into a single framework that links the theory to the four models. If payment rails are network goods (Katz & Shapiro, 1985; Rochet & Tirole, 2003), shocks that temporarily change the relative cost of cash can act as coordination devices that move adoption to a higher path (Crouzet et al., 2023), and the size and durability of that move depend on the infrastructure already in place (Alonso et al., 2023; Higgins, 2024). Diffusion theory then implies that growth should decelerate as the user base approaches saturation (Bass, 1969; Rogers, 2003). Because the shocks are few, externally dated and overlapping, the econometric literature cautions that break dates and unit-root inference are fragile (Perron, 1989; Andrews, 1993; Bai & Perron, 2003). This yields three empirical expectations, which we treat as interpretive guides rather than formally tested hypotheses.

First, if demonetisation acted as a coordination device, the Demonetisation window should show a change in the growth path that is not fully reversed afterwards (examined through M2 and M3; Sections 4.2 and 4.4). Second, if pandemic-era adoption built on an already-embedded network, the post-COVID period should show a shift in level or slope relative to the log-linear benchmark (examined through M3 and the Chow tests; Sections 4.2 and 4.3). Third, if adoption follows a diffusion curve, proportional growth should decelerate over the sample (examined through M4 and the annual series; Sections 3.2, 5 and 8). The design cannot separate these mechanisms from contemporaneous supply-side developments, a limitation developed in Sections 2.2 and 7, and the single-country series cannot test the cross-country complementarity discussed in Section 2.3.

3. Data and Methodology

3.1 Data

Our primary data are the NPCI UPI Monthly Statistics for April 2016 to April 2026, comprising 121 monthly observations. Variables used are: monthly transaction volume (millions), total transaction value (₹ crore), number of participating banks live on UPI, average transaction value (₹), and author-constructed binary indicator variables for the Demonetisation window (November 2016–March 2017), COVID Lockdown (March–June 2020), and Post-COVID period (July 2020–present).

Three months of zero recorded transaction volume (April-June 2016) are excluded from log-scale regressions because log(0) is undefined, yielding N=118 usable observations. This exclusion removes the platform's earliest pre-launch diffusion months from the sample: zero months are excluded because log(0) is undefined, and their omission means the regression results cannot speak to the platform's initial launch diffusion at all, not merely that any bias from omitting them is small; inference about that earliest stage should rely on the descriptive figures in Table 1 rather than on any regression estimate in this paper. Relatedly, the pre-demonetisation (n=4) and COVID-lockdown (n=4) eras, and to a lesser extent the demonetisation era (n=5), are short enough that their era-specific coefficients are highly unstable and should be read as

uncorrected proof

exploratory rather than as conventionally significant findings throughout this paper (Section 4.2). The April 2026 transaction value required manual correction from a non-standard comma format in the source file; the decade-cumulative and annualised transaction-value figures reported in the Abstract, Introduction, and Table 3 (Section 5) were verified against the source dataset.[1]

We define five growth eras using the authors' event-based regime definitions, anchored to publicly known policy and public-health dates rather than any NPCI-internal variable coding: Era 0 = pre-demonetisation (July-October 2016, n=4); Era 1 = demonetisation window (November 2016-March 2017, n=5); Era 2 = organic steady growth (April 2017-February 2020, n=35); Era 3 = COVID lockdown (March-June 2020, n=4); Era 4 = post-COVID (July 2020-April 2026, n=70). These boundaries are administrative and event-based rather than statistically estimated breakpoints. We treat these boundaries as maintained hypotheses evaluated using the a priori Chow tests, rolling-Chow search and multiple-break analysis of Section 3.3; CUSUM is used separately to assess overall parameter stability. We also flag that Era 0 and Era 3 contain only four monthly observations each, which limits the precision of any era-specific coefficient estimated from them (Section 4.2).

3.2 Econometric Models

All models are estimated in log-space, where the dependent variable is $\log(V_t)$, the natural logarithm of monthly transaction volume in millions. The logarithmic transformation converts multiplicative growth dynamics into an additive representation and facilitates interpretation of within-period growth rates. Heteroscedasticity-consistent (HC3) standard errors are used throughout.[2]

Model M1: Log-Linear Trend (Null Benchmark)

$$\log(V_t) = \alpha + \beta \cdot t + \varepsilon_t$$

This imposes a constant monthly growth rate across the full sample. The estimated $\hat{\beta}=0.076$ implies 7.9% monthly growth, a useful benchmark but one that ignores all structural variation. $R^2=0.769$, AIC=421.33.

Model M2: Log-Linear with Structural Level Shifts

$$\log(V_t) = \alpha + \beta \cdot t + \gamma_1 \cdot D_t + \gamma_2 \cdot PC_t + \varepsilon_t$$

[1] All data sourced from NPCI UPI Monthly Statistics (April 2016–April 2026). The April 2026 value was corrected from a non-standard comma format to ₹2,902,988.05 crore. Three months of zero transactions (April–June 2016) are excluded from log-scale regressions, yielding N=118.

[2] Heteroscedasticity-consistent (HC3) standard errors are used throughout. We also report Newey-West HAC(4) standard errors alongside HC3 (Section 4.2), since HC3 does not correct for the residual serial correlation documented in Section 4.5. These two corrections yield the same inference (significant vs. not, at conventional levels) for most M3 coefficients, but not all: the demonetisation-window level-shift is insignificant under HC3 and significant under HAC, the reverse of the usual pattern, and should be read with particular caution.

uncorrected proof

D_t is the Demonetisation dummy (1 for Nov 2016–Mar 2017) and PC_t is the Post-COVID dummy (1 from Jul 2020). This tests whether each shock produces a statistically significant level shift in log-volume above the global trend. R^2=0.800, AIC=408.55. The Post-COVID dummy is insignificant at conventional levels (t=0.41); the insignificant parameter may reflect the inability of M2 to represent simultaneous slope and level changes, since M2 allows a level shift but not an era-specific slope change at the same boundary, a limitation the piecewise structure of M3 is designed to address. We do not read this insignificance as evidence that COVID had no effect, only as evidence that M2's specification is not well suited to detecting one.

Model M3: Piecewise Slopes with Level Shifts

$$log(V_t) = \alpha + \Sigma_k \beta_k \cdot (t \cdot I_{kt}) + \delta_1 \cdot D_t + \delta_2 \cdot PC_t + \varepsilon_t$$

where I_{kt} is an indicator for era $k \in \{0,1,2,3,4\}$. This specification estimates a distinct monthly growth slope within each of the five eras, plus discrete level-shift dummies at the Demonetisation and Post-COVID transitions. It is the most flexible model and achieves R^2=0.977, Adj-R^2=0.975, AIC=162.66, BIC=184.83, RMSE=0.450, the strongest in-sample fit across all four specifications on these criteria. Model selection by AIC and BIC favours M3 over the log-linear benchmark (M1), though as Section 4.5 discusses, this comparison assumes approximately independent errors, an assumption violated here, so we treat it as supporting rather than decisive evidence. Because M3 includes both era-specific slopes and level-shift dummies at two of the four era boundaries, the level-shift and slope-change parameters are not fully orthogonal, particularly for the short pre-demonetisation (n=4) and demonetisation (n=5) eras; we flag the resulting imprecision explicitly where it is material (Section 4.2) rather than treat short-era coefficients as equally reliable as those from the longer eras.

Model M4: Quadratic Log-Trend

$$log(V_t) = \alpha + \beta_1 \cdot t + \beta_2 \cdot t^2 + \varepsilon_t$$

This specification tests for growth deceleration, one pattern consistent with, though not sufficient on its own to establish, S-curve platform maturation (Section 3.2 discusses this distinction further). The negative quadratic coefficient $\hat{\beta}_2$=−0.00110 (t=−9.39, p<0.001) indicates statistically significant deceleration in log-growth, with an implied turning point at approximately April 2024 (t≈93 months from July 2016). R^2=0.918, AIC=300.68. AIC favours M3 over M4, but M4 offers a more parsimonious description of the long-run growth arc; a single quadratic term cannot, however, distinguish genuine platform maturation from a temporary post-COVID demand normalisation, a distinction we return to in Section 8.[3]

3.3 Structural Break and Diagnostic Tests

[3] AIC = n·ln(RSS/n) + 2k; BIC = n·ln(RSS/n) + k·ln(n). Lower values indicate better relative in-sample fit after penalising for model complexity. M3 dominates on both criteria despite having 8 parameters, supporting, though not conclusively proving, that the piecewise structure fits better than the alternatives considered; AIC/BIC comparisons of this kind assume approximately independent errors, an assumption Section 4.5 shows is violated, so we treat this as supporting rather than confirming evidence.

Uncorrected Proof

We employ four diagnostic procedures. First, Chow (1960) structural break tests at each of the three a priori break dates (November 2016, March 2020, July 2020) using the baseline log-linear specification as the restricted model. F-statistics test whether the regression parameters differ significantly before and after each break.

Second, a rolling Chow test across all potential break dates identifies whether the maximum instability occurs at the hypothesised breaks or elsewhere. Third, CUSUM (cumulative sum of recursive residuals) analysis from Brown, Durbin, and Evans (1975) tests parameter stability over the full sample; a CUSUM process exceeding the ±1.36 critical boundary (5% level) indicates structural instability.

Fourth, autocorrelation diagnostics, the Durbin-Watson (1950) test statistic and the Ljung-Box (1978) Q(10) portmanteau test, assess whether M3 residuals exhibit serial correlation, and AR(1) fitting on residuals quantifies the persistence structure of any remaining autocorrelation.[4][5]

Two further methodological limitations bear on how these diagnostics should be read. First, as Section 4.5 documents, Durbin-Watson and Ljung-Box indicate substantial positive residual autocorrelation in M3; the HC3 standard errors reported throughout correct for heteroskedasticity but not for serial correlation. We report Newey-West HAC(4) standard errors alongside HC3 for this reason (Section 4.2); where autocorrelation materially widens a confidence interval or affects the validity of an F-test comparison, most notably the Chow tests reported against the log-linear baseline in Section 4.3, we flag this explicitly rather than treat the reported statistics as fully valid under classical assumptions. Second, we report formal unit-root tests on the log-volume series under both standard deterministic specifications, since the two can disagree. With an intercept only, an Augmented Dickey-Fuller (ADF) test fails to reject a unit root (statistic=-1.07, p=0.73, 10 lags selected by AIC); with an intercept and a deterministic trend, the ADF test rejects a unit root (statistic=-4.01, p=0.008, 10 lags). KPSS rejects both the level-stationary null (statistic=1.47, p=0.01) and the trend-stationary null (statistic=0.32, p=0.01), using automatic bandwidth selection (6 lags). Because the ADF conclusion reverses depending on the deterministic specification, unit-root results here are sensitive to trend and break specification rather than offering an unambiguous classification of the series, and we do not treat them as

Uncorrected proof

[4] The Durbin-Watson statistic of 0.552 for M3 residuals (well below the lower critical bound of ~1.6 for N=118) confirms positive serial autocorrelation. The estimated AR(1) coefficient of 0.694 indicates substantial persistence in residuals from one period to the next, consistent with momentum effects not fully captured by the piecewise structure. This residual autocorrelation does not automatically invalidate inference on the structural break dummies, which are identified by discrete regime transitions rather than smooth trends, but nor can it be assumed harmless. An attempted AR(1)-corrected sensitivity model (Section 4.2) did not converge to an economically interpretable solution, so we cannot use it to confirm or overturn inference on the level-shift dummies; we treat their robustness to serial correlation as an open question rather than resolved in either direction.

[5] The Chow (1960) test statistics reported here are point estimates at the three a priori defined break dates. As detailed in Section 4.3, the rolling Chow search in Figure 3(a) does not confirm this: its global maximum falls around January-April 2018, not at the three hypothesised dates, which rank 36th and 42nd of 106 candidate months. The candidate break dates are motivated by historical event timing; the a priori Chow tests evaluate those dates, while CUSUM assesses overall instability and does not identify break locations.

definitive proof of spurious regression for M1 or any other specification. An ADF test applied to M3's residuals rejects a unit root (statistic=-3.15, p=0.02), indicating that, once the three structural breaks are accounted for, the remaining deviations behave as a stationary process even though the undifferenced series does not unambiguously; we read this as consistent with, though not conclusive proof of, the piecewise approach describing the trend adequately. We note that the discrete-regime interpretation of the break dummies is likely more robust to this ambiguity than the precise numerical value of any single slope coefficient, and we treat point estimates accordingly, as indicative rather than exact, throughout the results that follow.

4. Empirical Results

4.1 Phase Descriptive Statistics

Table 1 summarises the five growth eras. The difference between the phases is not just in terms of growth rates, but in terms of the nature of the growth – the Demonetisation window had a spectacular proportional growth on a small base and the post-COVID window had a significant absolute growth over a protracted multi-year timeframe. Average transaction values declined from approximately ₹4,857 in the pre-Demonetisation period to approximately ₹1,299 in the Post-COVID period; Section 6 discusses this decline as a transaction-composition indicator rather than as direct evidence of demographic or income-based inclusion.

Uncorrected proof

Table 1: UPI Transaction Volume and Value by Growth Era (April 2016 – April 2026)

Phase	Period	Vol Start (M)	Vol End (M)	Phase Growth (%)	Avg Txn (₹)
Pre-Demonetisation	Jul–Oct 2016	0.09	0.1	~11%	~4,857
Demonetisation Window	Nov 2016–Mar 2017	0.29	6.37	~2,097%	~3,500
Organic / Steady Growth	Apr 2017–Feb 2020	7.2	1326	~18,312%	~1,720
COVID-19 Lockdown	Mar–Jun 2020	1247	1337	~7%	~1,778
Post-COVID Surge	Jul 2020–Apr 2026	1497	22347	~1,392%	~1,299

Source: NPCI UPI Monthly Statistics. Phase boundaries reflect the authors' event-based regime definitions (Section 3.1), not an NPCI-published variable coding. Vol. figures in millions. Phase Growth is (end-of-era volume / start-of-era volume - 1), expressed as a percentage. Avg Txn (₹) is the average transaction value in the final month of each era.

4.2 Model Estimates and Fit

The complete regression results are shown in Table 2. Heteroskedasticity-consistent (HC3) standard errors are reported throughout; significance is marked by conventional stars. These

corrections address heteroskedasticity but not the residual serial correlation documented in Section 4.5, which we discuss as a separate limitation on inference. A model comparison at the bottom of Table 2 shows M3 records the strongest in-sample OLS fit on R^2, AIC and BIC among the four specifications considered, though, as Section 4.5 discusses, these fit statistics should be read with the same autocorrelation caveat in mind.

Table 2: Regression Results for Models M1-M4 (Dependent variable: log monthly transaction volume; N=118 for M1, M2, M4; N=118 for M3; HC3-robust standard errors in parentheses; Newey-West HAC(4) standard errors in brackets, reported for all four models)

Parameter	Coefficient	HC3 SE	HAC(4) SE
M1: Log-Linear Trend			
Constant	2.639***	(0.442)	
t (monthly trend)	0.076***	(0.006)	[0.011]***
M2: Structural Dummy (Level Shifts Only)			
Constant	3.083***	(0.519)	
t (monthly trend)	0.069***	(0.009)	[0.016]***
D (Demonetisation level)	-2.712***	(0.736)	[0.758]***
PC (Post-COVID level)	0.148	(0.359)	[0.704]
M3: Piecewise Slopes with Level Shifts			
Constant	1.079†	(0.515)	[0.678]
t × Era0 (pre-Demonetisation, n=4)	-1.148*	(0.442)	[0.243]***
t × Era1 (Demonetisation, n=5)	0.697†	(0.409)	[0.128]***
t × Era2 (organic growth, n=35)	0.161***	(0.017)	[0.023]***
t × Era3 (COVID lockdown, n=4)	0.129***	(0.011)	[0.014]***
t × Era4 (Post-COVID, n=70)	0.038***	(0.001)	[0.002]***
D (Demonetisation level)	-5.101	(3.162)	[1.115]***
PC (Post-COVID level)	4.742***	(0.525)	[0.703]***
M4: Quadratic Log-Trend			
Constant	0.024	(0.508)	[0.911]
t	0.207***	(0.017)	[0.029]***
t^2	-0.0011***	(0.0001)	[0.0002]***

**** p<0.001, ** p<0.01, * p<0.05, † p<0.10. Parentheses report HC3-robust standard errors; brackets report Newey-West HAC(4) standard errors, since every model's residuals show significant serial correlation by Ljung-Box (see model-fit statistics below and Section 4.5). HAC(4) SEs are not reported for the M1 and M2 constants: none of this paper's claims rely on either constant, which in a trend specification reflects the extrapolated level at t=0 rather than a parameter of substantive interest, so we report the HC3 estimate for completeness and omit the additional HAC computation for these two cells. All four models are estimated by OLS on the same N=118 monthly observations (April-June 2016 excluded; Section 3.1) using one documented time index (t=1,...,118, July 2016-April 2026) and one set of era and event-date definitions (Section 3.1). An attempted AR(1)-corrected (iterative Cochrane-Orcutt) estimation*

uncorrected proof

of M3 converged numerically but produced economically uninterpretable coefficients (Section 4.2) and is not reported in this table.

Model fit statistics: M1 — R^2=0.769, Adj-R^2=0.767, AIC=421.33, BIC=426.87, RMSE=1.418, DW=0.029, Ljung-Box Q(10)=456.8 (p<0.001). M2 — R^2=0.800, Adj-R^2=0.795, AIC=408.55, BIC=419.63, RMSE=1.321, DW=0.130, Ljung-Box Q(10)=256.8 (p<0.001). M3 — R^2=0.977, Adj-R^2=0.975, AIC=162.66, BIC=184.83, RMSE=0.450, DW=0.552, Ljung-Box Q(10)=90.6 (p<0.001). M4 — R^2=0.918, Adj-R^2=0.917, AIC=300.68, BIC=309.00, RMSE=0.843, DW=0.069, Ljung-Box Q(10)=375.2 (p<0.001). AIC and BIC are computed identically across all four models (n·ln(RSS/n)+2k and n·ln(RSS/n)+k·ln(n) respectively) and are therefore directly comparable within this table: M3 records the lowest AIC/BIC and highest Adj-R^2 among the four specifications on these in-sample criteria. Every model's residuals show significant serial correlation by Ljung-Box, which is why HAC(4) standard errors are reported alongside HC3 for all four models in this table.

Source: Authors' calculations based on NPCI data.

We flag one coefficient in Table 2 as too imprecisely estimated to interpret substantively: the pre-demonetisation era slope in M3 (β=-1.148) is estimated from only four monthly observations and carries a standard error more than a third the size of the point estimate. The coefficient is statistically significant under both reported covariance estimators, but because it is estimated from only four observations, we do not interpret it substantively. We report it for completeness but do not draw on it elsewhere in the paper. The raw monthly series are plotted in Figure 1 with shading to denote the different regimes, while the log-scale series in Figure 2 has fitted values of M1, M3, and M4 overlaid. Both the piecewise and quadratic models track the observed series more closely than the log-linear model around the demonetisation and post-COVID periods, where the log-linear model's misfit is visually apparent; this is a description of relative fit and is not, on its own, evidence for any particular causal account of why the series deviates from the log-linear path.

uncorrected proof

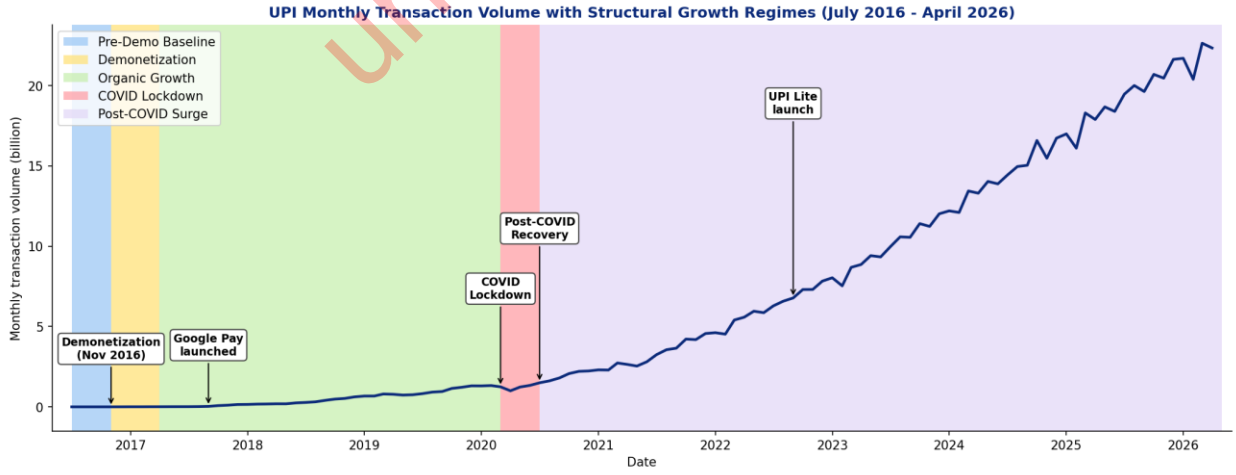
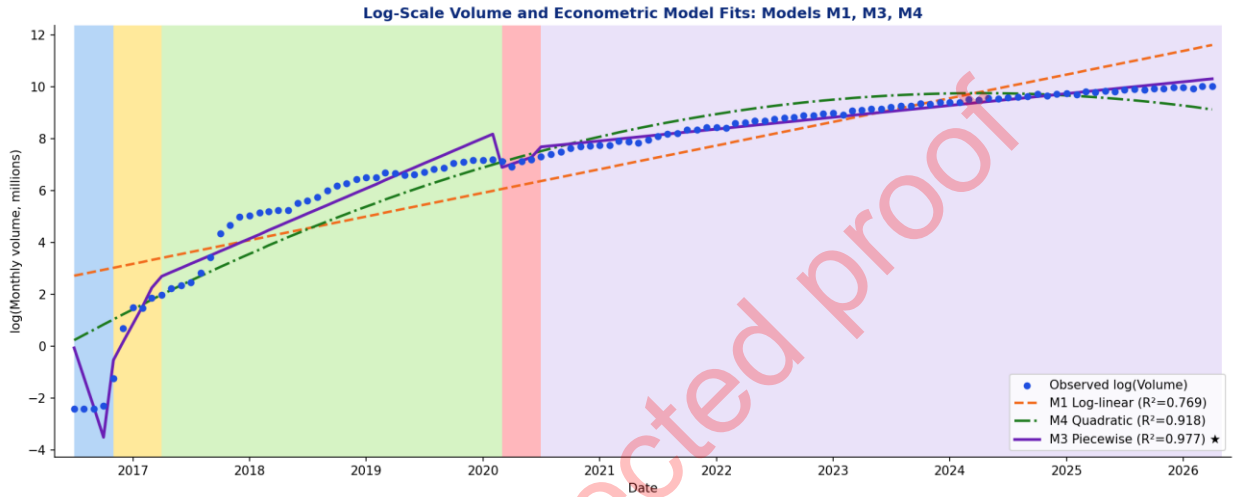


Figure 1: UPI Monthly Transaction Volume with Structural Growth Regimes (April 2016 – April 2026). Shaded regions correspond to the five analytically defined eras

Source: NPCI.

Figure 1 the figure plots the raw monthly volume series from July 2016 to April 2026 on a linear scale, with background shading marking the five eras defined in Section 3.1 (pre-demonetisation, demonetisation window, organic growth, COVID lockdown and post-COVID). The annotations mark selected ecosystem events for orientation only and are not estimated break dates. On this scale the series stays close to the horizontal axis until about 2020, dips briefly at the lockdown, and rises steeply thereafter with visible month-to-month variation, which is why the logarithmic view in Figure 2 is used for model comparison.



uncorrected proof

Figure 2: Log-Scale Monthly Volume with Fitted Values from Models M1 (log-linear), M3 (piecewise, best-fitting on AIC/BIC), and M4 (quadratic). April-June 2016 are excluded from this and all other logarithmic plots in this paper because log(0) is undefined (Section 3.1). Shading as in Figure 1.

Source: Authors' calculations based on NPCI data.

Figure 2 observed log volume is shown as points, with fitted values from M1 as a dashed straight line, from M4 as a dash-dotted curve and from M3 as a solid piecewise line. The M1 line lies above the observed series in the earliest period, below it through most of 2018 to 2020, and above it again from about 2024, which is the visual misfit of a constant growth rate. The M4 curve rises concavely, peaks around 2024 and then turns down, so that it falls below the observed series at the end of the sample. The M3 line tracks the post-2020 observations closely, but it departs from the data at the very start of the sample, where the eras contain only four or five observations, and it overshoots the observed series at the end of the organic-growth era before dropping at the lockdown boundary.

4.3 Structural Break Tests

The Chow tests indicate high levels of statistical significance for parameter breaks at all three hypothesised dates. The largest F-statistic is obtained for the COVID-onset break (March 2020, $p<0.001$), followed by the post-COVID recovery (July 2020, $F=125.1$) and demonetisation (November 2016, $F=63.1$). As discussed in Section 3.3, these F-statistics are computed under the classical assumption of serially uncorrelated errors; given the autocorrelation documented in Section 4.5, the reported p-values likely overstate the precision of this comparison. CUSUM independently confirms general parameter instability in M1, but does not identify or corroborate the three prespecified dates; the rolling Chow search, discussed next, does not provide that corroboration either and requires a more cautious reading.

We re-ran the rolling Chow search directly on the source data: a single-break, two-segment log-linear split scanned against every candidate month with at least six observations on each side. The search does not peak at the three a priori dates. The three dates remain individually significant ($F=63.1$, 167.2, 125.1), but the search's global maximum ($F\approx700$-800) falls around January-April 2018, in the organic-growth era, and the two testable a priori dates rank 36th and 42nd of 106 candidate months. A single-break scan is a coarse tool for validating breaks in a process we model as having five regimes, so this does not overturn the a priori Chow and CUSUM evidence, but it does mean the rolling search does not independently corroborate our date choices, and the confirmation-bias risk (Section 3.1) is not resolved by it. As a further check, we ran an unrestricted multiple-break search: dynamic-programming exact optimisation of a piecewise-linear cost function (Truong, Oudre & Vayatis, 2020, implemented in the Python 'ruptures' package), allowing breaks in both intercept and slope, with a minimum segment length of 6 months and no maximum imposed. We did not apply a formal information-criterion penalty to select the number of breaks, since a preliminary BIC search favoured an implausibly large number (BIC continued to fall through at least 6 breaks); we instead report results for $k=4$ breaks, chosen only for illustrative comparison against our three hypothesised dates, and flag the number of breaks as a modelling choice we have not resolved rather than one selected by a formal criterion. At $k=4$, the estimated break months are December 2016, February 2020, May 2022, and May 2024; the first two fall approximately one month from the Demonetisation and COVID-onset dates respectively, the third does not correspond to any date this paper otherwise discusses, and the fourth falls near the M4 turning point (Section 3.2). The Post-COVID date (July 2020) is not separately recovered, likely because it falls within the minimum-segment-length window of the COVID-onset break. We did not compute confidence intervals for these estimated break dates. We read this as partial, not confirmatory, corroboration of our a priori dates, and flag the choice of k and the absence of confidence intervals as limitations of this check rather than a fully resolved robustness test. The maximum CUSUM statistic of 2.51 is well above the 5 percent critical value (1.36), indicating that the parameters of the baseline log-linear specification are unstable over the sample. Consistent with Brown, Durbin, and Evans (1975), this establishes that some form of regime change is present; it does not identify the location,

Uncorrected proof

number, or cause of any break, does not by itself identify which alternative specification, among M2-M4, best describes it, and we do not treat it as confirming the piecewise model, or the three prespecified dates, specifically. Figure 3(b) plots the growth path implied by the absence of structural breaks (M1) against actual log-volume; the shaded area is the resulting gap in the post-COVID period, read here as a description of the deviation from a linear counterfactual rather than as a causally estimated COVID effect.

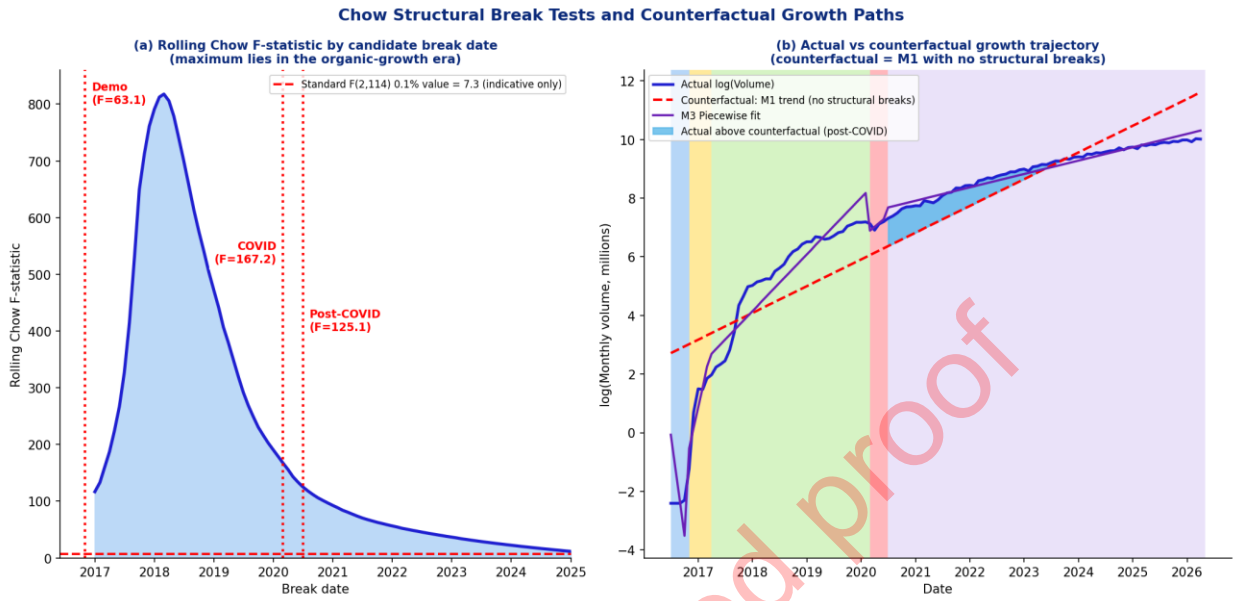


Figure 3: (a) Rolling Chow F-Statistics across all candidate break months; the three vertical markers show the a priori hypothesised dates (Demonetisation, COVID onset, Post-COVID recovery), which are individually significant but do not correspond to the global maximum of the search (Section 4.3). (b) Counterfactual growth path (M1, no breaks) vs. actual log-volume, with the shaded area showing the difference from the fitted M1 benchmark

Source: Authors' calculations based on NPCI data.

Figure 3 panel (a) plots the rolling Chow F-statistic against each candidate break date, with vertical markers at the three a priori dates and their F-statistics (63.1, 167.2 and 125.1 in the text). The curve rises to a maximum in early 2018, in the organic-growth era, and declines steadily towards later dates, so the global maximum does not coincide with any hypothesised date. The Demonetisation marker lies to the left of the plotted range because the search requires at least six observations on each side of a candidate break (Section 4.3). Panel (b) plots actual log volume against the M1 no-break trend (dashed) and the M3 fit; the shaded area marks where actual volume exceeds the M1 counterfactual after mid-2020. The gap is widest around the July 2020 transition and narrows over the following years until the M1 line overtakes the actual series

uncorrected proof

in the mid-2020s, which is the pattern described as a substantial but non-permanent shift in Section 7.2.

4.4 Era-Level Growth Rate Decomposition

M3's era-specific slope estimates show markedly different monthly growth rates across phases. The organic steady-growth period (April 2017-February 2020), the only era in our sample without an identified exogenous shock, has the highest sustained slope at +0.161 log-units/month, or 17.5 percent monthly compound growth, over 35 consecutive months. We read this as consistent with, though not conclusive proof of, sustained supply-side ecosystem growth during this window. The post-COVID period shows a smaller proportional slope (+0.038, or 3.9 percent per month), but because it applies to a far larger transaction base, the implied absolute monthly increase in transactions is substantially larger than in the steady-growth era; we note that comparing percentage growth rates across eras of very different absolute size, as we do throughout this paper, requires care, and we flag this distinction wherever it could otherwise be read as implying a slowdown in absolute terms.

The slope in the Demonetisation window was large and positive following the demonetisation shock (+0.697 log-units/month, or 100(e^0.697-1)≈101% per month using the standard log-to-percentage conversion), while the change in levels was negative (δ_1=-5.101), consistent with volumes being pushed down from a near-zero base and then recovering quickly; we describe this as a data pattern and do not attach a specific behavioural mechanism (e.g., "forced trial" or "habit formation") to it, since the four- and five-observation eras involved are too short to distinguish between competing behavioural accounts (Section 3.1).

4.5 Model Diagnostics

Here is a four panel diagnostic suite for Model M3 (Figure 4). The time-series residuals from panel (a) are coloured by era; residuals visibly cluster around certain periods, particularly near the break dates, rather than scattering evenly across the sample, which is a visual signature of the serial correlation quantified below. The Q-Q plot (residual-quantile correlation r=0.950) is shown in panel (b) and points towards moderate fat tails at the extreme quantiles and nearly normal residuals otherwise. The ACF for M3 residuals with 95% confidence bounds are plotted in panel (c), autocorrelation is present at lags 1-4, and DW=0.552 indicates positive serial correlation (AR(1) coefficient $\hat{\varphi}$=0.694); the Ljung-Box Q(10) test rejects the null of no autocorrelation up to 10 lags at p<0.001 (Q(10)=90.6). The CUSUM on M1 residuals in Panel (d) crosses its critical boundary within the sample, indicating parameter instability in M1 overall; consistent with Section 4.3, this establishes instability but does not by itself identify the exact break dates, so the visual proximity of the crossing to the demonetisation and post-COVID windows in this panel should not be read as the test independently locating those dates.

Uncorrected proof

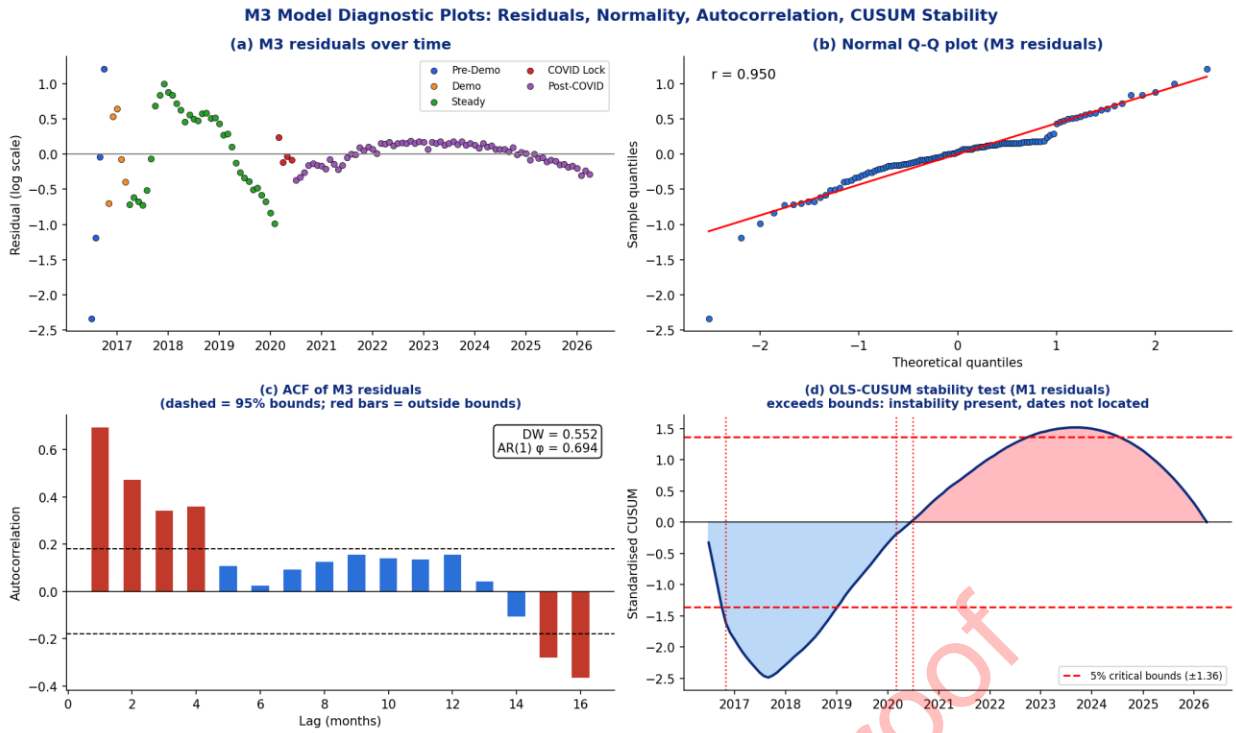


Figure 4: Four-panel diagnostic suite for Model M3. (a) Residuals over time by era. (b) Normal Q-Q plot. (c) ACF of residuals with DW and AR(1) statistics. (d) CUSUM stability test on M1 residuals with 5% critical bounds

Source: Authors' calculations based on NPCI data.

Figure 4 panel (a) shows M3 residuals over time, coloured by era; the residuals move in long runs of the same sign, notably through the organic-growth era, and form a smooth hump in the post-COVID period, rather than scattering around zero, which is the visual signature of serial correlation. Panel (b) is a normal Q-Q plot in which the points follow the reference line over the central range and depart from it in both tails, with one large negative residual from the earliest months. Panel (c) shows the residual autocorrelation function, which is positive and decays with the lag, with the shortest lags outside the 95% bounds. Panel (d) plots the standardised CUSUM path from the M1 residuals, which leaves the 5% critical bounds, indicating parameter instability in M1 without locating any particular break date (Section 4.3).

The autocorrelation in M3 residuals ($\hat{\varphi}$=0.694, DW=0.552) is a genuine weakness of the specification: it indicates the piecewise structure does not fully remove month-to-month momentum, and it means the reported HC3 standard errors, t-statistics, and even the AIC-based model comparisons in Section 4.2 understate the true uncertainty around M3's coefficients, potentially substantially. We do not treat this as a minor caveat. ARMA-augmented or GLS-corrected estimation would be the appropriate remedy in a full econometric treatment and is a natural extension of this work. The direction of the identified breaks is corroborated

Uncorrected proof

independently by the a priori Chow tests, though not by the rolling Chow search (Section 4.3); CUSUM confirms general parameter instability but does not corroborate the specific direction or timing of individual breaks. These results are unlikely to be pure artifacts of the autocorrelation, but the precise magnitude of any individual coefficient, including the level-shift and slope estimates discussed throughout Sections 4 through 9, should be read as indicative rather than exact, with wider true confidence intervals than the tables report.

5. Annual Growth Trajectory and Financial Year Analysis

The financial year summary is given in Table 3. The 5,024% YoY growth in FY2017-18 reflects a mathematical base effect, since total volume in FY2016-17 was only 17.86 million transactions. The acceleration from 78% YoY growth in FY2020-21 to 106% in FY2021-22 is broadly consistent with, though not an independent confirmation of, the level-shift result from Model M3 (Section 4); the two draw on the same underlying series and so cannot be treated as separate evidence. This path is shown in Figure 5, which plots the percentage-growth slowdown alongside the absolute volume increase.

Table 3: UPI Annual Summary by Financial Year (FY2016-17 to FY2025-26)

Financial Year	Total Volume (Mn)	Total Value (₹ Cr)	YoY Vol Growth (%)	Avg Txn (₹)
2016-17	17.86	6952	—	3893
2017-18	915	1,09,832	5,024%	1200
2018-19	5392	8,76,971	489%	1627
2019-20	12519	21,31,730	132%	1703
2020-21	22331	41,03,654	78%	1838
2021-22	45968	84,17,572	106%	1831
2022-23	83751	1,39,20,676	82%	1662
2023-24	1,31,130	1,99,95,086	57%	1525
2024-25	1,85,866	2,60,56,955	42%	1402
2025-26	2,41,617	3,14,23,250	30%	1301

Source: NPCI UPI FY Summary data. Financial years follow the standard Indian convention (April-March); FY2025-26 covers April 2025-March 2026, verified by summing the underlying monthly series (241,617 million transactions, ₹3,14,23,250 crore), which matches the totals reported in this table. April 2026 itself falls in FY2026-27 and is included in the monthly series (Table 1, Section 4) but not in any FY2025-26 total above. YoY growth rates are calculated from total annual volume.

Uncorrected proof

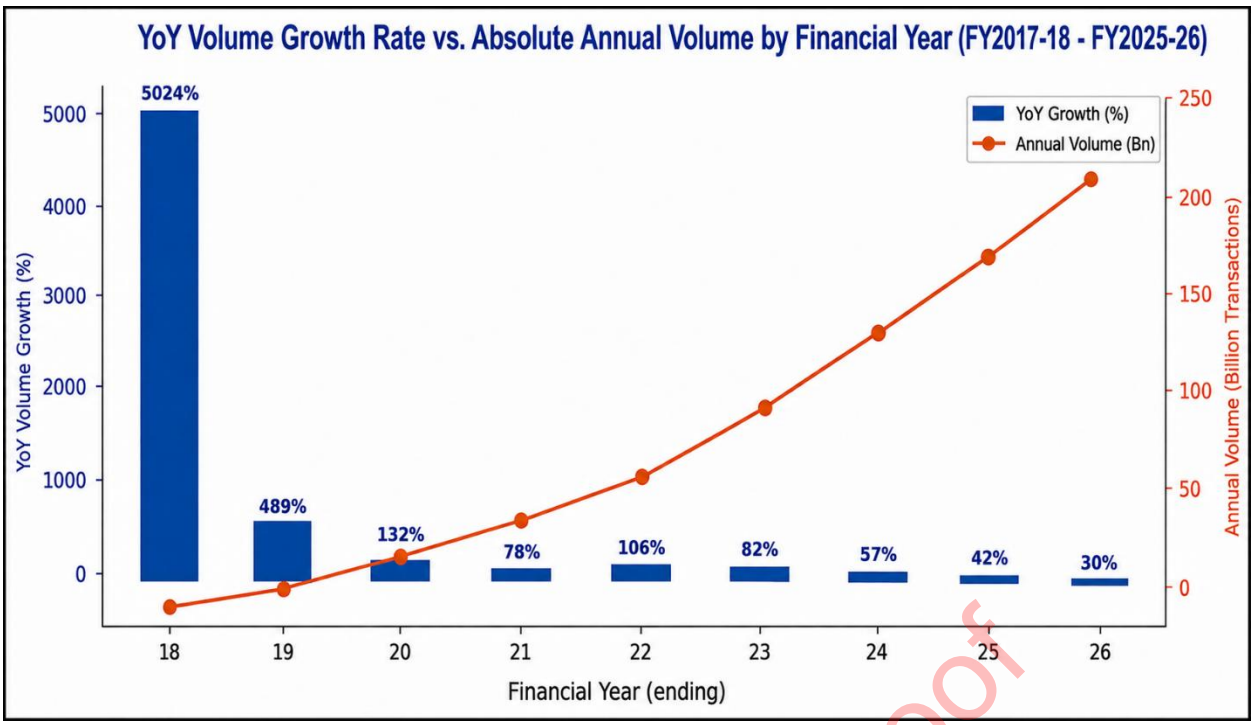


Figure 5: YoY Volume Growth Rate (bars, left axis) and Total Annual Volume in Billions (line, right axis) by Financial Year, FY2017-18 to FY2025-26. Deceleration in growth rate is accompanied by accelerating absolute volume addition.

Source: NPCI.

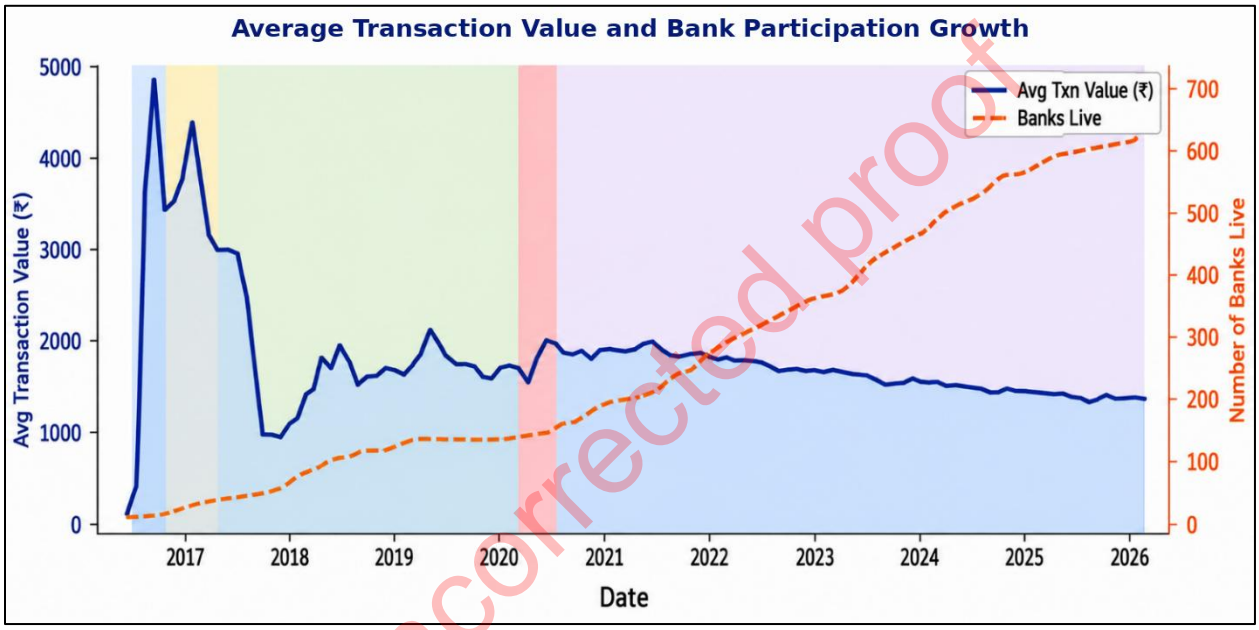
In percentage-growth space, the deceleration from 5,024% to 30% over eight financial years traces a roughly log-linear path, broadly consistent with the deceleration M4 estimates in log space (Section 3.2). This percentage deceleration is a mechanical consequence of a growing base and should not, on its own, be read as evidence against a slowdown or for one: the absolute volume added in FY2025-26, approximately 55,751 million transactions, exceeds the platform's entire cumulative volume through FY2021-22, which is consistent with continued strong absolute growth, but a declining percentage rate is also what a genuinely maturing platform would show. We treat the percentage and absolute series as complementary descriptions rather than let either one settle the maturation-versus-slowdown question, which we return to in Section 8.

One commonly used transaction-composition indicator associated with the Digital India financial-inclusion mandate is the average transaction value column, which fell from ₹3,893 in FY2016-17 to ₹1,301 in FY2025-26. We treat this as a transaction-composition indicator rather than a direct measure, for reasons discussed in Section 6. Rural-urban convergence in UPI transaction density is commonly discussed as beginning around FY2021-22, coinciding with the period when the average transaction value in our series first fell consistently below ₹1,600, though we do not claim the two series jointly establish causation.

uncorrected proof

6. Transaction Composition and the Financial-Inclusion Question

Before presenting these figures, we state the caveat that governs how they should be read. Average transaction value is, at best, a transaction-composition indicator, not a direct measure of financial-inclusion depth. A falling average transaction value is consistent with the entry of new low-income or rural users, but it is equally consistent with existing users transacting more frequently in smaller amounts (for example, frequent small merchant payments crowding out larger transfers in the average), or with the withdrawal of MDR subsidies on certain transaction categories over the sample period. NPCI's published series does not decompose average transaction value by user segment, transaction type, or subsidy regime, so this paper cannot separate these channels, and we treat every claim in this section as suggestive rather than demonstrated. With that caveat in place: Figure 6 plots average transaction value alongside bank-participation growth. The average fell from ₹4,857 in October 2016 to ₹1,299 by April 2026, a 73% decrease.



Uncorrected proof

Figure 6: Average Transaction Value (₹, left axis) as a transaction-composition indicator (see caveat above), and Banks Live on UPI (right axis), July 2016 - April 2026. The long-run value decline is consistent with, but does not on its own establish, demographic deepening toward lower-income users.

Source: NPCI

Average transaction value was initially high and volatile (₹917-₹4,857 during the pre-Demonetisation and Demonetisation window, July 2016-March 2017) and subsequently declined, reaching ₹1,678 by early 2020 and continuing to decline through FY2022-23. This pattern is consistent with a greater relative frequency of smaller transactions over time, but the aggregate NPCI series cannot determine whether the decline reflects entry by new users, changes

in the mix of person-to-person and merchant payments, greater transaction frequency among existing users, changes in subsidy or pricing arrangements, or other compositional factors. We therefore treat the series as a transaction-composition indicator rather than as direct evidence of demographic or income-based inclusion, and do not attribute the acceleration in decline after FY2021-22 to a specific user group without disaggregated data we do not have.

This pattern reflects a broader monetisation tension: broadening reach toward lower-value users runs against revenue models built on Merchant Discount Rates (MDR) or interchange income, since a falling average transaction value implies falling per-transaction revenue even as transaction counts reach new highs. We return to the sustainability question this raises in Section 9.3.

As stated at the outset of this section, we flag decomposing average transaction value into its constituent channels (new-user entry, transaction-mix shift, and subsidy changes) as a priority for future work using more granular, non-public NPCI data.

7. Comparative Interpretive Assessment

Table 4 synthesises the era-level evidence from Tables 1-3 and Sections 4-6, read alongside the literature reviewed in Section 2, into a comparative summary across three candidate explanations. We are explicit about what this comparison is and is not: it is a descriptive, interpretive synthesis, not a formally identified causal decomposition, since none of Models M1-M4 include a direct measure of digital-policy intensity, and the demonetisation and COVID windows are not independently identified from the contemporaneous supply-side developments discussed in Section 2.2. Read this way, we do not find the evidence consistent with a single, monocausal driver; we read it as more consistent with a sequenced complementarity, in which each candidate factor plays a distinct role, while acknowledging that our methods cannot rule out alternative sequencings.

Uncorrected proof

Table 4: Comparative Interpretive Summary of Three Candidate Explanations: Demonetisation, COVID-19, and Digital India

Dimension	Demonetisation	COVID-19	Digital India / Policy
Nature of shock	Acute, coercive demand-side	Exogenous health/mobility crisis	Sustained public infrastructure
Timing	Immediate (Nov–Dec 2016)	Gradual from Jul 2020	Continuous 2014–present
Phase volume growth	+2,097% in 5 months	+1,392% over 70 months	Continuous across all phases (descriptive; not separately estimated)
Chow F-statistic	63.1 ($p<0.001$)	125.1 ($p<0.001$)	Not represented by a separate estimated parameter

M3 level shift	δ=−5.101 (below trend)	δ=+4.742*** (positive OLS level parameter; robustness to serial correlation inconclusive, Section 4.2)	N/A
Reversal after shock?	Partial (Feb 2017 dip)	No reversal observed during the sample period	Not applicable
Avg txn value signal	₹3,500 (higher average transaction value during the event window)	₹1,778 (stable)	₹1,301 (declining average transaction value)
Bank participation	30→44 banks	148→216 banks	21→713 banks (full arc)
Long-run interpretation	Short-window increase from a very low transaction base	Persistent post-2020 adoption shift (robustness to serial correlation unresolved, §4.2)	Candidate enabling condition (interpretive, not econometrically identified; §7)

M3 statistics: δ_1 = Demonetisation level shift, δ_2 = Post-COVID level shift. Chow F at point estimates for the three a priori breaks.

Source: Authors' calculations based on NPCI data.

uncorrected proof

7.1 Demonetisation-Period Growth Pattern

The Chow test F-statistic of 63.1 ($p<0.001$) indicates a statistically significant parameter break coinciding with demonetisation, subject to the autocorrelation caveat discussed in Section 4.3. During the November 2016-March 2017 event window, the estimated within-period slope is positive while the associated level parameter is negative (Table 2). Because this period contains only five observations and overlaps with several contemporaneous developments (Section 2.2), we report these estimates descriptively and do not assign them a behavioural interpretation.

The demonetisation window contributed only about 17.49 million transactions, a small fraction of the platform's cumulative decade volume, with a higher average transaction value (roughly ₹3,500) than the eras before or after it (Table 1). We report this as an observable pattern in the data rather than an inference about who the early users were, since the aggregate NPCI series does not identify user income or location. Volumes did not fully revert after the window closed, but nor did they sustain a rate above the organic steady-era growth that follows (Section 4.4); we describe this as neither full reversal nor full retention, without attributing it to a specific mechanism.

7.2 COVID-19 and the Post-2020 Adoption Shift

Across all four models, the largest structural parameter is the post-COVID level shift in M3 (δ_2=+4.742, significant under both HC3 (SE=0.525) and HAC(4) (SE=0.703) inference), though "most accurately estimated" should be read relative to the other coefficients in this paper rather than as a claim of high absolute precision, given the autocorrelation discussed in Section 4.5. The Post-COVID dummy coefficient cannot be read in isolation as a fixed multiplier on the M1 log-linear counterfactual, since the piecewise specification simultaneously changes the active era-specific slope: comparing the two models' full fitted trajectories (Section 4.3) shows the gap between them is largest at the July 2020 transition itself and narrows substantially over the following years, closing and then reversing by the mid-2020s as M1's steeper global trend catches up. We therefore describe this as a substantial but non-permanent shift in trajectory, not as a fixed-size effect. Among the three prespecified Chow tests, the COVID-onset statistic (F=167.2) is the largest, consistent with this being the most pronounced of the three hypothesised structural breaks; this is a different claim from saying it is the largest value in the rolling search (Section 4.3), which it is not.

One plausible reading is that COVID-19 reduced practical and behavioural barriers to digital payments rather than creating demand from nothing: as of March 2020, UPI already handled 1,247 million transactions per month, so the pandemic's role may have been to erode merchant resistance to digital terminals, hygiene concerns around cash, and institutional hesitancy toward digital payment, rather than to manufacture demand from nothing. We favour this reading as consistent with the pattern in our data, while noting it is drawn from the literature rather than independently established here. The post-July 2020 within-era slope of 3.9 percent per month has not reverted across the 70 months of post-COVID data in our sample, which we describe as a persistent rather than strictly permanent shift, since our sample cannot rule out that a future shock or saturation effect could alter this path.

7.3 Digital India: A Candidate Enabling Condition

The strongest evidence consistent with a Digital India account is the 35-month organic growth period (April 2017-February 2020): volume growth of +18,312 percent at a month-on-month rate of 17.5 percent, without an identified exogenous shock in our framework. This period coincides with the entry of Google Pay, the expansion of QR-code acceptance, the rollout of UPI 2.0 features, and merchant formalisation following GST implementation, developments consistent with ecosystem momentum building independently of Demonetisation or COVID, though we again note that this era's slope, like the others, is a within-era average and does not itself identify which of these concurrent developments mattered most.

Bank participation increased substantially over the sample, from 21 to 713 institutions. This is consistent with supply-side ecosystem expansion, but the present model does not establish whether bank participation caused transaction growth, followed transaction growth, or evolved jointly with it; we do not attribute a lag direction to this relationship.

uncorrected proof

If the effects of demand-side shocks depend partly on pre-existing supply-side infrastructure, some portion of the observed post-COVID change may reflect interaction with accumulated digital infrastructure. The present single-country design cannot test this possibility separately.

8. Growth Trajectory: Saturation or Maturation?

If the quadratic log-trend in Model M4 is an accurate description of the underlying process, log-growth decelerates toward a turning point (vertex of the fitted quadratic) around April 2024. In the diffusion-of-innovations framework of Rogers (2003), declining proportional growth can be consistent with movement toward later stages of adoption. However, the quadratic turning point estimated here does not by itself establish an S-shaped diffusion process; as noted in Section 3.2, a single quadratic coefficient cannot by itself distinguish genuine S-curve maturation from a temporary post-COVID demand normalisation, and we treat both as live possibilities in what follows.

India's UPI transaction rate is approximately 195 transactions per person per year (TPPY) as of April 2026, calculated directly from the annualised volume in this paper's own data and a standard population estimate. We do not compare this figure to other countries' payment systems, since we could not source comparator TPPY figures with confidence, and cross-country payment-volume comparisons are in any case sensitive to differences in measurement and market structure that this paper is not positioned to adjudicate. A further potential growth segment is rural and semi-urban India, where PMJDY (Jan Dhan) account penetration is in the hundreds of millions with general account operability commonly reported in the 74-81 percent range; whether active UPI usage specifically lags general account activity is plausible but not a claim we can support with a verified figure, so we note the possibility without attaching a number to it.

Beyond smartphone access, two structural constraints on further inclusion, connectivity and smartphone dependency, have been partly addressed by NPCI's UPI 123PAY (for feature phones, launched December 2022) and UPI Lite (an offline payment framework). Separately, cross-border UPI corridors are a further potential growth channel. Published reporting indicates cross-border UPI transactions grew roughly 20-fold year-on-year to approximately 755,000 in FY2024-25, a far smaller base than domestic volumes and one this paper does not attempt to project forward.

A single point forecast for December 2027 depends heavily on which in-sample pattern is assumed to continue; we report three deterministic mechanical extrapolations rather than a statistical forecast with uncertainty intervals, since none of Models M1-M4 was designed to produce out-of-sample prediction intervals.

Table 5: Illustrative December 2027 Volume Scenarios Under Alternative Growth Assumptions

uncorrected proof

Scenario	Model	Assumption	Dec 2027 monthly volume (million)	Caveat
M3 continuation	Piecewise (M3)	Post-COVID within-era slope (3.9%/month) continues unchanged	~63,900	Mechanical upper scenario; assumes no further deceleration
Recent actual trend	Trailing 12-month growth	Growth rate observed April 2025-April 2026 continues	~32,400	Sensitive to the 12-month window chosen
M4 extrapolation	Quadratic (M4)	Fitted quadratic trend extended 20 months beyond the sample	~2,100	Implausible: falls below the April 2026 actual level, illustrating that quadratic extrapolation is unreliable once the fitted curve passes its vertex

The scenarios span roughly a 30-fold range (~2,100 to ~63,900 million transactions per month), which we treat as the substantive finding: any single December 2027 figure is highly sensitive to modelling choices the data cannot adjudicate between. We do not endorse any one scenario as more likely than the others and do not report a single headline number in the main text for this reason.

Uncorrected proof

9. Policy Implications

9.1 The Complementarity Thesis

The discussion that follows synthesises the descriptive evidence above into a policy narrative; given the identification limitations noted throughout this paper (Sections 2-4, 7), we present it as our reading of the balance of evidence rather than as a proven causal chain, and we would characterise it as a sequential-complementarity interpretation rather than a finding. In this reading, the evidence is consistent with the possibility that pre-existing Digital India infrastructure enhanced the capacity of the Demonetisation and COVID-19 shocks to accelerate adoption; Demonetisation appears to have provided an initial trigger that moved the platform off a near-zero base; and COVID-19 coincides with a persistent change in the observed transaction-growth trajectory. We hold this narrative more tentatively than a single sentence can convey, and note again that our models cannot separately identify each of these roles.

The UPI case motivates a testable hypothesis for comparative research: pre-existing interoperable payment infrastructure may increase the capacity of subsequent demand-side events to accelerate digital-payment adoption. Evaluating this proposition requires cross-country data and an explicit identification strategy; the present single-country study cannot establish its general applicability. UPI's JAM Trinity foundations predate the platform itself rather than having been added to it (World Bank, 2023), which is one reason we emphasise sequencing in this discussion.

9.2 Open-Stack Design as Force Multiplier

UPI's institutional design is open, interoperable and bank-based, with private applications competing on a shared payment infrastructure. This institutional architecture is one plausible contributor to the expansion in bank participation and application-provider diversity, although the present analysis does not estimate its independent causal contribution.

9.3 Monetisation and Sustainability

A pressing policy question concerns the long-run commercial sustainability of a payment ecosystem operating under a zero-MDR regime. The declining average transaction value (₹1,299 by April 2026) is consistent with a shift toward smaller-value transactions, although it does not independently establish financial inclusion. The development of credit-linked UPI products is one step toward reconciling inclusion outcomes with platform sustainability. We note, however, that MDR on UPI P2M transactions has remained at zero throughout our sample period, including for high-value transactions, so any sustainability gain in this channel currently depends on government incentive payments and other mechanisms outside the merchant fee structure itself; we do not have direct evidence in this paper on whether credit-linked products or other mechanisms can realistically sustain the ecosystem at scale, and flag this as an important question for future work rather than one this paper resolves. How this balance is managed will likely bear on whether the post-COVID growth path documented here continues.

10. Conclusion

This paper applies a suite of time-series models to NPCI's 121-month UPI series to document the timing and magnitude of structural breaks in India's digital-payments growth path, and to discuss, with explicit caveats about causal identification, how three candidate explanations align with that pattern. The key empirical findings are:

First, Model M3 (piecewise slopes and level shifts) fits the data substantially better than a log-linear benchmark on standard information criteria (R^2=0.977, AIC=162.66 versus 421.33 for M1), though its residuals show pronounced positive serial correlation that widens the true uncertainty around its coefficients beyond what the reported HC3 statistics show (Section 4.5). Second, the prespecified Chow tests at the three a priori break dates produce large classical F-statistics (63.1, 167.2, and 125.1, all $p<0.001$), while CUSUM independently rejects overall

uncorrected proof

parameter stability in the log-linear benchmark (max|CUSUM|=2.51 vs. critical 1.36); CUSUM does not identify or validate the three prespecified dates specifically. A rolling Chow search does not corroborate the three dates either: its global maximum falls elsewhere in the sample (Section 4.3), which strengthens rather than resolves the confirmation-bias concern raised about our a priori approach. These tests establish that instability is present; they do not by themselves establish which specification or which underlying cause best explains it. Third, the estimated post-COVID level shift (δ_2=+4.742, significant under both HC3 and HAC(4) inference) is the largest OLS level parameter among the structural breaks estimated in Model M3. A full fitted-trajectory comparison (Section 4.3) shows the gap between M3 and the log-linear benchmark is not constant: it is largest at the July 2020 transition and narrows, then reverses, over the following years as the log-linear model's steeper global trend catches up. We interpret this as a substantial but non-permanent break in trajectory rather than a fixed or lasting multiplier, and note that an attempted AR(1)-corrected sensitivity model (iterative Cochrane-Orcutt) converged numerically to $\rho \approx 0.96$ but produced economically uninterpretable coefficients (Section 4.2), so the robustness of this level shift to serial correlation remains an open question rather than resolved in either direction. Fourth, a quadratic specification (M4, R^2=0.918) implies a declining fitted marginal growth rate over the sample and reaches its turning point around April 2024; this should not be read as showing that deceleration began only in April 2024, and is a pattern consistent with, though not conclusive evidence of, S-curve-style maturation.

Read alongside the literature reviewed in Section 2, these results are consistent with a sequenced narrative in which pre-existing digital infrastructure may have enhanced the capacity of demonetisation and COVID-19 to accelerate adoption, and COVID-19 coincided with a persistent change in the observed transaction-growth trajectory. We want to be explicit about what this paper's methods can and cannot establish: Chow tests and piecewise trend models identify when and how much a growth path changed, not, on their own, why. Several developments discussed in this paper, including Jio's 4G rollout, the BHIM launch, and GST-driven merchant formalisation, occurred close in time to the events we study and are not separately identified in our specifications. The organic-growth era (April 2017-February 2020), in which volumes grew 18,312 percent without an identified exogenous shock in our framework, is the single piece of evidence in this paper most consistent with a Digital India account, but it is consistency, not proof; the interpretation offered here is best read as a plausible, literature-consistent narrative rather than a causally identified decomposition, and we would resist describing any of the three forces as strictly necessary on the strength of this evidence alone.

Uncorrected proof

Limitations

Five limitations bear directly on how these results should be used. First, the three break dates are externally specified from known policy events rather than statistically detected, which raises a risk of confirmation bias. An unrestricted multiple-break search (Section 4.3) recovers two of the three dates only approximately (within one to two months) and does not recover the third as a separately identified break; we treat this as partial, not confirmatory, evidence and do not

consider the confirmation-bias concern resolved. Second, M3's residuals exhibit substantial positive autocorrelation (Durbin-Watson=0.552, AR(1) phi=0.694 in our replication). We report Newey-West HAC(4) standard errors alongside HC3 throughout (Section 4.2): these are larger than HC3 for most coefficients and, for the Demonetisation level-shift specifically, change which correction shows significance. We additionally attempted an AR(1)-corrected model using iterative Cochrane-Orcutt estimation (100 iterations, convergence tolerance 1e-8, final N=117); the algorithm converged numerically to $\rho \approx 0.96$, but the resulting coefficients describe the data poorly (negative R^2 in the original units), so we do not report specific AR(1)-corrected coefficient values or treat them as a robustness finding. The Cochrane-Orcutt sensitivity exercise converged numerically but did not produce an economically interpretable transformed specification. We therefore treat it as an inconclusive robustness check. The robustness of the Post-COVID level shift to serial correlation remains an open question. Third, we report formal stationarity tests (ADF and KPSS on log-volume; Section 3.3) under both standard deterministic specifications, since they disagree: with an intercept only, ADF fails to reject a unit root (statistic=-1.07, p=0.73); with an intercept and trend, ADF rejects a unit root (statistic=-4.01, p=0.008). KPSS rejects both the level-stationary and trend-stationary nulls (p=0.01 in each case). Because the ADF conclusion reverses with the deterministic specification, we treat this as a genuine sensitivity rather than unambiguous evidence of spurious regression. An ADF test on M3's own residuals rejects a unit root, consistent with, though not conclusively establishing, the piecewise structure describing the trend adequately. Fourth, none of the four models includes a direct measure of Digital India policy intensity or a cross-country comparison group, so the comparative assessment in Section 7 and the cross-country extension in Section 9 are interpretive rather than econometrically identified causal claims. Fifth, several era-specific coefficients, particularly those for the four-observation pre-demonetisation era and the four-observation COVID-lockdown era, are estimated from too few observations to support conventional inference; we treat these as exploratory rather than as confirmed findings throughout (Section 4.2).

The UPI case motivates the hypothesis that interoperable, government-backed payment infrastructure may strengthen the capacity of payment systems to respond to demand-side shocks. Testing whether such gains are more persistent requires comparative evidence beyond the present study.

uncorrected proof

References

1. Agarwal, S., Ghosh, P., Li, J., & Ruan, T. (2024). Digital payments and consumption: Evidence from the 2016 demonetization in India. The Review of Financial Studies, 37(8), 2550–2585. https://doi.org/10.1093/rfs/hhae005
2. Alonso, C., Bhojwani, T., Hanedar, E., Prihardini, D., Uña, G., & Zhabska, K. (2023). Stacking up the benefits: Lessons from India's digital journey (IMF Working Paper No. 2023/078). International Monetary Fund. https://doi.org/10.5089/9798400240416.001

3. Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. Econometrica, 61(4), 821–856.
4. Auer, R., Cornelli, G., & Frost, J. (2020). Covid-19, cash, and the future of payments (BIS Bulletin No. 3). Bank for International Settlements.
5. Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. Journal of Applied Econometrics, 18(1), 1–22. https://doi.org/10.1002/jae.659
6. Bass, F. M. (1969). A new product growth for model consumer durables. Management Science, 15(5), 215–227.
7. Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society, Series B, 37(2), 149–192.
8. Chodorow-Reich, G., Gopinath, G., Mishra, P., & Narayanan, A. (2020). Cash and the economy: Evidence from India's demonetization. The Quarterly Journal of Economics, 135(1), 57–103. https://doi.org/10.1093/qje/qjz027
9. Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica, 28(3), 591–605.
10. Crouzet, N., Gupta, A., & Mezzanotti, F. (2023). Shocks and technology adoption: Evidence from electronic payment systems. Journal of Political Economy, 131(11). https://doi.org/10.1086/724847
11. Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression. Biometrika, 37(3–4), 409–428.
12. Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. Econometrica, 25(4), 501–522.
13. Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in U.S. labor productivity. Journal of Economic Perspectives, 15(4), 117–128.
14. Higgins, S. (2024). Financial technology adoption: Network externalities of cashless payments in Mexico. American Economic Review, 114(11).
15. Jack, W., & Suri, T. (2014). Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. American Economic Review, 104(1), 183–223.
16. Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. American Economic Review, 75(3), 424–440.
17. Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303.
18. Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2016). Building state capacity: Evidence from biometric smartcards in India. American Economic Review, 106(10), 2895–2929.
19. National Payments Corporation of India. (2016-2026). UPI product statistics. https://www.npci.org.in/what-we-do/upi/upi-ecosystem-statistics
20. Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3), 703–708.

Uncorrected proof

21. Parker, G., & Van Alstyne, M. (2005). Two-sided network effects: A theory of information product design. Management Science, 51(10), 1494–1504.
22. Perron, P. (1989). The Great Crash, the oil price shock, and the unit root hypothesis. Econometrica, 57(6), 1361–1401.
23. Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. Journal of the European Economic Association, 1(4), 990–1029.
24. Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.
25. Sahay, R., von Allmen, U. E., Lahreche, A., Khera, P., Ogawa, S., Bazarbash, M., & Beaton, K. (2020). The promise of fintech: Financial inclusion in the post COVID-19 era. IMF Departmental Paper No. 2020/009.
26. Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. Signal Processing, 167, 107299.
27. World Bank. (2023). G20 policy recommendations for advancing financial inclusion and productivity gains through digital public infrastructure: Global Partnership for Financial Inclusion 2023. World Bank. http://hdl.handle.net/10986/40421

uncorrected proof